

Models of Human Movement Detection in Video Frames Using Machine Learning Technology

Arwa Darwish Alzughaibi

University of Technology Sydney
Faculty of Engineering and Information Technology

Supervisor

Dr. Zenon Chaczko

This dissertation is submitted for the degree of Doctor of Philosophy

Autumn 2019

Certificate of Original Authorship

I, Arwa Alzughaibi declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:

Signature: Signature removed prior to publication.

Date: 25-10-2019

Abstract

In recent years, human movement detection has been a very active and vibrant field of research. Examples of successful applications that adopt discoveries and developments in the human movement detection domain include pedestrian detection, intelligent monitoring systems and activity recognition. Thus far, much research work has been done and many different approaches explored to optimise accuracy, performance and productivity of human detection system. However, there are still many open questions related to these issues remain unresolved. The following study aims to develop a more effective human detection system that is able to operate in various environmental conditions and application contexts including illumination changes, pose and scale differences. The existing solutions for human detection and tracking systems often do not produce reliable and consistent results without considering changes in environmental conditions. In this work, a novel illumination invariant human detection algorithm is proposed which applies an alternative approach for the selection of orientation extraction and texture extraction features to identify human shapes in various illumination and contrast invariant conditions. An innovative human detection approach is also proposed to resolve and improve results of pose invariant cases. This research work involves the exploration of feature extraction techniques that offer superior results when dealing with human subjects in pose invariant conditions. Another innovation of this investigation is the design of a human detection and tracking model that can work in situations where human subjects are occluded within frames. In the proposed models, several pre-processing and post-processing stages are used for reducing detection errors and to improve the model performance. These approaches help to classify the frame contents more efficiently. The proposed computational solutions are extensively tested and performance evaluated using the standard datasets. The resulting output is encouraging when compared to the reported and the State-Of-The-Art human detection algorithms. The newly developed methods are tested using two practical applications and are included in this thesis as action research studies. In the first action study, children activity monitoring system is built to test the human movement detection algorithms, whilst the second action study involves a construction of an expert system for counting humans in a moving crowd to validate the effectiveness of the proposal computational models. These two key action research studies report high feasibility and viability of the proposed solutions

إهداء

أهدي ثمرة جهدي

إلى من قاسمني الحياه بيسرها وعسرها

إلى من تحلو الحياه بوجودهم

إلى من هم لحياتي حياة .. إلى قرة عيني ونبض قلبي

رنيم ♥ فرح ♥ إبراهيم

Acknowledgments

In the name of Allah, the Most Gracious and the Most Merciful

Alhamdulillah, I praise and thank Allah SWT for his greatness and for giving me the strength, knowledge, ability, and courage to complete this thesis. Without his blessings, this work could not exist.

First and foremost, I offer my special appreciation to Dr. Zenon Chaczko, my research supervisor for providing his heartfelt support, inspiration, guidance and for his invaluable help of constructive comments and suggestions at all times. I would also like to thank Dr. Chris Chiu who helped to review and proofread this thesis, as well as to my fellow friends at the UTS Faculty of Engineering and Information Technology for their support throughout my studies.

On the personal front, I would like to express my sincerest gratitude to my parents, husband, daughters, son and sister Mashael, for their love, sincere prayers and encouragement throughout my study.

Then, to those who supported me throughout my journey to achieving my goal, your kindness means a lot to me. Thanks for being there when my family couldn't.

Finally, my acknowledgment would be incomplete without thanking and expressing my deepest gratitude to the Saudi Arabia Culture Mission and Taibah University for providing a chance and funding to engage this research.

Contents

Abstract	ii
Acknowledgements	iv
I Research Background and Problem Investigation	1
1 Introduction to Human Movement Detection	2
1.1 Human Detection Systems: An Overview	3
1.2 Research Motivation	4
1.3 Research Gap	5
1.4 Research Aims	6
1.4.1 Research Objectives	7
1.5 Research Hypothesis	7
1.5.1 Hypothesis Statement	8
1.6 List of Publications	8
1.7 Structure of the Thesis	9
2 The Challenge of Human Movement Detection	11
2.1 Introduction	11
2.2 Human Movement Detection Approaches	15
2.2.1 Supervised Techniques	16
2.2.2 Unsupervised Techniques	36
2.2.3 Comparison of Detection Techniques	39
2.3 Human Tracking Systems	40
2.3.1 2D Tracking Approaches	40
2.3.2 3D Tracking Approaches	41
2.4 Datasets for Human Detection	43
2.5 Human Movement Detection Systems: Factors for Performance Evaluation	44
2.5.1 Accuracy	44
2.5.2 Computational Time	47
2.5.3 Memory Requirements	47
2.6 Discussion	50
2.7 Conclusion	53

3	Methodology and Mathematical Apparatus	54
3.1	Introduction	54
3.2	Research Design	54
3.3	Research methods	55
3.3.1	Feature extraction methods	55
3.3.2	Machine learning Methods	60
3.3.3	Enhancement Techniques in Human detection	63
3.3.4	Research Analysis	71
3.4	Conclusion	73
II	Research Contributions and Problem Solutions	74
4	Illumination and Pose Invariant Human Detection Models	75
4.1	The Problem	75
4.2	The Proposed Models	77
4.2.1	Illumination Invariant Human Detection Model	77
4.2.2	Pose Invariant Human Detection Model	79
4.2.3	Datasets	80
4.2.4	Training Phase	82
4.2.5	Testing Phase	85
4.2.6	Pseudocode	90
4.3	Experimental Results and Discussion	93
4.4	Conclusion	105
5	Enhancing the Precision of Human Movement Detection	107
5.1	The Problem	107
5.2	The Proposed Model	108
5.2.1	Datasets	109
5.2.2	Training Phase	110
5.2.3	Testing Phase	111
5.2.4	Pseudo Code	113
5.3	Experimental Results and Discussion	115
5.4	Conclusion	120

6	An Improved Person Tracking System with Occlusion Detection	121
6.1	The Problem	121
6.2	The Proposed Model	121
6.2.1	Datasets	122
6.2.2	Training	123
6.2.3	Testing	124
6.2.4	Pseudocodes	128
6.2.5	Experimental Results and Discussion	131
6.3	Conclusion	135
7	Action Studies and Application	137
7.1	Introduction	137
7.2	Action Study 1: Early Warning System for Monitoring Child Activity	137
7.2.1	Prior Art and Their Limitations	138
7.2.2	Aim of the Study: Kids Activity Monitoring System	139
7.2.3	Scope of the Study	139
7.2.4	Design and Implementation	140
7.2.5	Discussion	148
7.3	Action Study 2: Expert System for Counting Humans in the Moving Crowd	149
7.3.1	Aim of the Case Study	149
7.3.2	System Design and Architecture	149
7.3.3	Implementation	153
7.3.4	Discussion	154
7.4	Conclusion	157

8	Conclusion and Future Directions	158
8.1	Overview	158
8.2	Research Findings	158
8.2.1	Improved efficiency in detecting human subjects in images with poor contrast	158
8.2.2	Detection efficiency in case of pose and illumination vari- ant human subjects	159
8.2.3	Precision enhancement using multimodal feature extraction	159
8.2.4	Occlusion detection and correction in unconstrained video frames	159
8.2.5	Smart surveillance application for monitoring kids	160
8.2.6	Crowd analytics by counting human subjects	160
8.3	Research Contribution	160
8.4	Future Directions	161
III	Bibliography and Appendix	163
	References	164
9	Appendix	182

List of Figures

1.1	Generic architecture of a human detection system.	3
1.2	Efficiency of state-of-the-art techniques on INRIA data-set.	6
1.3	Mind map of research.	10
2.1	Sample output images of human detection systems	11
2.2	Still shots from videos showing change in poses of a single human subject	13
2.3	Example image frames with illumination variations	13
2.4	Examples of partially occluded pedestrians.	14
2.5	Samples of appearance and occlusion change coupled together. . .	15
2.6	Basic architecture of a machine learning based human detection .	17
2.7	(a) Input image, (b) Edges calculated for the input image	21
2.8	(a) Input Image, (b) Gradient image, (c) Orientation of each gradi .	22
2.9	The Haar-like binary basis functions to detect specific features . .	24
2.10	The process of obtaining local binary patterns. The original image	25
2.11	Background subtraction calculates the foreground mask by	27
2.12	Background subtraction approach where (a) is the background . .	28
2.13	(a) Shows the left image, (b) shows the generated depth map, and	28
2.14	Optical flow vectors overlaid onto the original video frames,	29
2.15	Top row row visualises two consecutive framing from video capture	30
2.16	Detection of human body parts using the part models.	34
2.17	The drastic change in the structure of human body parts	34
2.18	Multi-person pose estimating. Body parts belonging to the same .	35
2.19	Common failure cases: (a) Rare appearance or pose, (b) Missing parts	35

2.20	Examples of different types of occlusion i.e. due to another	37
2.21	Detection accuracy of various approaches at (a) non-occluded, . .	38
2.22	Sample images from publicly available datasets.	43
2.23	Overview: best results on the Caltech-USA pedestrian.	50
2.24	Progress of human detection systems from 2004 to 2014.	50
3.1	The process of obtaining local binary patterns.	56
3.2	Thresholding in LBP	57
3.3	A simple classification model	59
3.4	Different types of classification model from training samples. . .	60
3.5	Overview of ACF detector.	62
3.6	Median filtering.	63
3.7	(a) Input image with salt and pepper noise, (b) Normal median . .	64
3.8	(a) Input images, (b) Enhanced image using Unsharp masking . .	66
3.9	(a) Input image (b) Gaussian filtered image	66
3.10	(a) RGB model of colour, (b) CMY model of colour, (c) HSI model of colour	68
3.11	Search space pruning. (a) Input image, (b) Result image	69
3.12	Search space pruning. (a) Input image with multiple detections . .	70
4.1	A few video frames with different level of illumination (source: UCF sports action dataset and change detection 2014 dataset). . . .	75
4.2	A few video frames with different postures (source: Poses In The Wild dataset).	76
4.3	Block diagram of the illumination invariant human model	78
4.4	Block representation of the proposed pose invariant human	80
4.5	A few sample image obtained by INRIA database.	81
4.6	A few frames from UCF sports action dataset.	82
4.7	A few sample video frames from change detection 2014 dataset. .	82
4.8	Few video frames from Poses In The Wild dataset.	82
4.9	HOG vectors and its normalized LPQ codeword histogram.	83

4.10 HOG features.	84
4.11 LBP features.	84
4.12 Optical flow vectors based on the amount of difference	85
4.13 Search space pruning allows to decrease the search space	85
4.14 Slide window of 70×134 size is run for each test image (applied using UCF sports action dataset sample).	87
4.15 Multi-scale representation of the image.	87
4.16 Several bounding boxes detected around the same human subject (applied using UCF sports action dataset sample).	88
4.17 Application of non-maximal suppression to eliminate the	88
4.18 Sliding window of size 64×128 is run over each frame (applied using Poses In The Wild dataset sample).	89
4.19 Execution phases of proposed illumination invariant human	93
4.20 Execution phases of proposal illumination invariant human	94
4.21 A number of sample detection results made by the proposed illu- mination	94
4.22 A number of sample detection results made by the proposed illu- mination	94
4.23 Performance of detection by propositioned illumination invariant human	95
4.24 Performance of detection by propositioned illumination invariant human	96
4.25 Miss rate vs FPR ROC of different human detectors	97
4.26 Miss rate vs FPR ROC of different human detectors	97
4.27 TPR vs FPR ROC of different human detectors	98
4.28 TPR vs FPR ROC of different human detectors	98
4.29 Detection results of the proposed pose invariant human detection	99
4.30 Some examples of detections made by different detectors	100
4.31 Some examples of detections made by different detectors	101
4.32 Few examples of detections made by different detectors	102
4.33 Some sample results of detections made by different detectors	103
4.34 Miss rate vs FPR ROC of different human detectors	104

4.35	TPR vs FPR ROC of different human detectors	104
5.1	Block level representation of propositioned model of human de- tection	108
5.2	Sample images from INRIA person dataset.	109
5.3	Sample frames from CDW 2014 dataset sequences.	110
5.4	(a) Sample RGB image (b) HOG feature map (c) HSV channel. . . .	110
5.5	Several bounding boxes detected around the same human subject (applied using CDW 2014 dataset sample).	112
5.6	Application of non-maximised suppressives	112
5.7	Implementation steps of the precision enhanced human detector (applied using CDW 2014 dataset sample).	115
5.8	Implementation steps of the proposed precision enhanced human	116
5.9	Detections by the proposed detector on a frame sequence	116
5.10	Few examples of detections made by different detectors	117
5.11	Few examples of detections made by different detectors	118
5.12	Precision-Recall ROC of different humanoid detectors	119
5.13	TPR-FPR ROC plot of different humanoid detectors	119
6.1	Block diagram of the proposed method.	122
6.2	(a)(b) Image sampling from high resolution (INRIA) Dataset	123
6.3	Detailed Flow chart of the Proposed method.	125
6.4	Performance curves in Caltech Dataset	132
6.5	Results while using (a) Caltech,(b) TUD stadtmittle	135
7.1	Workflow of the Kid's Activity Monitoring System.	141
7.2	KAMS Conceptual Architecture.	143
7.3	KAMS Execution Architecture.	144
7.4	KAMS Solution Architecture.	145
7.5	(a) Shows the login screen for the registered users	146
7.6	shows sample status update regarding the activities of the kids. .	147
7.7	(a) Depicts results of detection of kid in the video	147
7.8	(a) Depicts sample date for kids profile	148
7.9	Workflow of the Human Counting System.	151
7.10	Sample test results (applied using TUD pedestrian dataset sample).	153
7.11	Sample results from human counting system (applied using UCSD and TUD pedestrian dataset sample).	156

List of Tables

2.1	Details of the modified implementations of histogram of gradients.	23
2.2	Details of the modified implementations of LBP.	26
2.3	A generic comparison of object detection techniques with respect .	39
2.4	Datasets for applications relating to humans	42
2.5	Comparison of state-of-the art human detection and tracking systems, table adapted from (Solichin et al 2014)	48
2.6	Effect of view angle on detection performance.	52
2.7	Effect of lighting condition on detection performance.	52
2.8	Effect of human density on detection performance	52
4.1	Comparison of proposed illumination invariant human detector .	99
4.2	Comparison of proposed pose invariant human detector	105
5.1	Comparison of proposed precision enhanced human detector . . .	120
6.1	Matric values in each confidence score.	134
7.1	Performance statistics in different density test frames, where respectively L, M, H in 3rd column is Low, Medium and High. . . .	155

List of Algorithms

3.1	ACF Person Detection	65
4.1	Search Space Pruning	90
4.2	Multi-Scale Detection	91
4.3	Training pose invariant human detector	92
4.4	Testing pose invariant human detector	92
5.1	Training Phase	113
5.2	Testing Phase	114
6.1	Pre-processing	128
6.2	Training	129
6.3	Testing	130
6.4	Occlusion Detection	131
6.5	ACF Person Detection	131

Nomenclature

AHE	Adaptive Histogram Equalization
AR	Augmented Reality
AUC	Area Under the Curve
CPU	Central Processing Unit
CSLBP	Centre Symmetric LBP
DPM	Deformable Part Model
EKF	Extended Kalman Filter
FN	False Negative
FNR	False Negative Rate
FP	False Positive
FPpI	False Positive per Image
FPR	False Positive Rate
HCI	Human Computer Interaction
LBP	Local Binary Pattern
LDCF	Local Decorrelated Channel Features
LTP	Local Ternary Pattern
MIL	Multiple Instance Learning
MoG	Mixture of Gaussian
MR	Miss Rate
NRLBP	Non-Redundant Local Binary Patterns
OF	Optical Flow
PSF	Point Spread Function
ROC	Receiver Operator Characteristics
TN	True Negative
TP	True Positive
TPR	True Positive Rate

Terms and Definitions

Accuracy	Refers specifically to the closeness of the measurement to the true value.
Bounding Box	The kind of method of annotation to mark a region or object of interest from a source image. In a two-dimensional image, it is typically representative of the coordinate value of the top-left pixels, the height of the box and width of the box.
Classification	The process of image classification on the premise of recognising characteristics or features within an image.
Computer Vision	People use their brains and eyes to visualise the world that they perceive, thus computer vision is the computer science aiming to provide a similar, or optimal ability to a computer or machine. Computer vision deals with the automated extracting, analysing and understanding of meaningful data from an image or image sequence. It deals with the design of an algorithmic and theoretical construct to accomplish the automatic understanding of visuals.
Confidence Score	Defining the event probability, or input probability to fall in differing classes. If a class has a great probability then confidence is high. The value of confidence is calculated for one input, and providing meaning as to the algorithm's confidence for the class.
Detection	The process of determining and finding notable objects in an image.
Feature Extraction	The process of detecting and extracting the most representing data from a raw source as features, then deriving discovered features from the original such as to lower the cost of measuring the features, improve efficiency of the classifier and enable classification with higher accuracy.
Feature Representation	The process of providing a uniquely representative point for every sequence of video based on features that are extracted.
Ground Truth	The term used to describe the data offered by observing directly, that is empirical evidence, instead of the information made by inferences.

Human Detection	The task of localisation of human instances that are present within an image, and it is mostly done by finding all the locations in the image, for all available scales, and making a comparison in a small region at every location with well-known people patterns or templates.
Precision	Refers specifically to the closeness of the measurement to each other.
Receiver Operating Characteristic Curve	It is a curve represented as a graphic plot chart, depicting the diagnosis of a binary classifier system, because of the variation of its discriminatory threshold.